FUNDAÇÃO GETULIO VARGAS
EPGE
Escola de Pós-Graduação em Economia

Ensaios Econômicos

Escola de
Pós Graduação
em Economia
da Fundação
Getulio Vargas

Nº 642 ISSN 0104-8910

A Panel Data Approach to Economic Forecasting: The Bias–Corrected Average Forecast

João Victor Issler, Luiz Renato Lima

Janeiro de 2007

Os artigos publicados são de inteira responsabilidade de seus autores. As opiniões neles emitidas não exprimem, necessariamente, o ponto de vista da Fundação Getulio Vargas.

A Panel Data Approach to Economic Forecasting: The Bias-Corrected Average Forecast*

João Victor Issler
Luiz Renato Lima†
Graduate School of Economics – EPGE
Getulio Vargas Foundation
email: jissler@fgv.br and luizr@fgv.br

First Draft: December, 2006.

Abstract

In this paper, we propose a novel approach to econometric forecasting of stationary and ergodic time series within a panel-data framework. Our key element is to employ the bias-corrected average forecast. Using panel-data sequential asymptotics we show that it is potentially superior to other techniques in several contexts. In particular,

*We are especially grateful to Raffaella Giacomini, Clive Granger, Marcelo Moreira, Zhijie Xiao, and Hal White, for their suggestions regarding the initial idea of the paper. We also benefited from comments given by Marcelo Fernandes, Marcelo Medeiros, and from the participants of the conference "Econometrics in Rio." We thank Claudia Rodrigues for research assistance and gratefully acknowledge support given by CNPq-Brazil, CAPES, and Pronex. João Victor Issler thanks the hospitality of the Rady School of Management, and the Department of Economics of UCSD, where parts of this paper were written. The usual disclaimer applies.

†Corresponding Author.

it delivers a zero-limiting mean-squared error if the number of forecasts and the number of post-sample time periods is sufficiently large. We also develop a zero-mean test for the average bias. Monte-Carlo simulations are conducted to evaluate the performance of this new technique in finite samples. An empirical exercise, based upon data from well known surveys is also presented. Overall, these results show promise for the bias-corrected average forecast.

Keywords: Panel-Data Econometrics, Pooling of Forecasts, Forecast Combination Puzzle, Common Features.
J.E.L. Codes:C14, C32, C33, C53, G11.

1 Introduction

Bates and Granger(1969) made the econometric profession aware of the benefits of forecast combination when a limited number of forecasts is considered. The widespread use of different combination techniques has lead to an interesting puzzle from the econometrics point of view – the well known *forecast combination puzzle*: if we consider a fixed number of forecasts ($N < \infty$), combining them using equal weights ($1/N$) fare better than using "optimal weights" constructed to outperform any other forecast combination.

Regardless of how one combine forecasts, if the series being forecast is stationary and ergodic, and there is enough diversification among forecasts, we should expect that a weak law-of-large-numbers (WLLN) applies to well-behaved forecast combinations. Indeed, Timmermann(2006) uses financial-economic arguments based upon risk diversification to defend the idea of pooling of forecasts. This motivates labeling it "a financial approach to economic forecasts," since it is based on a principle so keen on finance; see, e.g., Ross (1976), Chamberlain and Rothschild (1983), and Connor and Korajzcyk (1986, 1993). Of course, to obtain this WLLN result, the number of forecasts has to diverge ($N \to \infty$), which entails the use of asymptotic panel-data techniques. In our view, one of the reasons why pooling forecasts has not yet been given a full asymptotic treatment is that forecasting is frequently thought to be a time-series experiment, not a panel-data experiment. As far as we know, despite its obvious benefits, there has been no work where the pooling forecasts was considered in a panel-data context, with the number of forecasts (N) and time-series observations (T) diverging without bounds.

In this paper, we propose a novel approach to econometric forecasting of stationary and ergodic series within a panel-data framework. First, we decompose individual forecasts into three components: the series being forecast, a time-invariant forecast bias, and a zero-mean forecast error. We show that the series being forecast is a *common feature* of all individual forecasts; see Engle and Kozicki(1993). Second, when $N, T \to \infty$, and we use standard tools from panel-data asymptotic theory, we show that the pooling of forecasts delivers optimal limiting forecasts in the sense that they have a zero mean-squared error. The key element of this result is the use of the *bias-corrected average forecast* – equal weights in combining forecasts coupled with a bias-correction term. The use of equal weights avoids estimating forecast weights, which contributes to reduce forecast variance, although potentially at the cost of an increase in bias. The use of a bias-correction term eliminates any possible detrimental effect arising from equal weighting. One important element of our technique is to use the forecast combination puzzle to our advantage, but now in an asymptotic context.

The use of the bias-corrected average forecast is a parsimonious choice in forecasting that delivers optimal forecasts in a mean-squared error sense – zero limiting mean-squared error. The only parameter we need to estimate is the mean bias, which requires the use of the sequential asymptotic approach developed by Phillips and Moon (1999). Indeed, the only way we could increase parsimony in our framework is by doing without any bias correction. To test the usefulness of performing bias correction, we developed a zero-mean test for the average bias which draws upon the work of Conley (1999) on random fields.

Despite the lack of panel-data work on the pooling of forecasts, there has been panel-data research on forecasting focusing on pooling of information; see Stock and Watson (1999 and 2002a and b) and Forni et al. (2000, 2003). The former is related to forecast combination and operates a reduction on the space of forecasts. The latter operates a reduction on a set of highly correlated regressors. In principle, forecasting can benefit from the use of both procedures. However, the payoff of pooling forecasts is greater than that of pooling information: while pooling information delivers optimal forecasts in the mean-squared error sense (Stock and Watson), it cannot drive the mean-squared forecast error to zero as the pooling of forecasts can.

One important element of our technique is the introduction of a bias-correction term. If a WLLN applies to a equal-weight forecast combination, we cannot guarantee a non-zero mean-squared error in forecasting, since the limit average bias of all forecasts may be non-zero. In this context, one interesting question that can be asked is the following: why are forecasts biased? From an economic standpoint, Laster, Bennett and Geoum (1999) show that professional forecasters behave strategically (i.e., they bias forecasts) if their payoffs depend mostly on publicity from the forecasts than from forecast-accuracy itself. Since one way to generate publicity is to deviate from a consensus (average) forecast, rewarding publicity may induce bias. From an econometric point of view, Patton and Timmermann (2006) consider an additional reason for the existence of bias in forecasts: what may look like forecast bias under a specific loss function may be just the consequence of the forecaster using a different loss function in producing the forecast[1]. Hoogstrate, Palm and Pfann (2000) show that pooling cross-sectional slopes can help in forecasting. One of the potential reasons why this procedure works in practice is that only cross-sectional slopes are pooled, not individual effects, showing that the latter may be working as a bias-correction device. A final reason for bias in forecasts is non-stationarity of the variable being forecast or of a subset of the conditioning variables. This is explored by Hendry and

[1] Also, Clements and Hendry's (1999) work on intercept correction can be viewed as a study of bias.

Clements (2002) and Clements and Hendry (2006).

Given that important forecast studies are motivated by bias in forecasting, it seems desirable to build a forecasting device that incorporates bias correction. We view the introduction of the bias-corrected average forecast as one of the original contributions of this paper. The way we estimate the bias-correction term relies on the use of a forecast-specific component to capture the bias in individual forecasts. Of course, this can only be fully studied asymptotically within a panel-data framework, which reinforces our initial choice of approach.

The ideas in this paper are related to research done in two different fields. From econometrics, it is related to the common-features literature after Engle and Kozicki (1993). Indeed, we attempt to bridge the gap between a large literature on common features applied to macroeconomics, e.g., Vahid and Engle (1993, 1997), Engle and Issler(1995), Issler and Vahid (2001, 2006) and Vahid and Issler (2002), and the econometrics literature on forecasting related to common factors, forecast combination, bias correction, and structural breaks, perhaps best represented by the work of Bates and Granger (1969), Granger and Ramanathan(1984), Forni et al. (2000, 2003), Hendry and Clements (2002), Stock and Watson (2002a and b), Elliott and Timmermann (2003, 2004, 2005), and, more recently, by the excellent surveys of Clements and Hendry (2006), Stock and Watson (2006), and Timmermann (2006) – all contained in Elliott, Granger and Timmermann (2006). From finance and econometrics, our approach is related to the work on factor analysis when the number of assets is large, to recent work on panel-data asymptotics, and to panel-data methods focusing on financial applications, perhaps best exemplified by the work of Ross (1976), Chamberlain and Rothschild (1983), Connor and Korajzcyk (1986, 1993), Phillips and Moon (1999), Bai and Ng (2002, 2004), Bai (2005), and Pesaran (2005), and Araujo, Issler and Fernandes (2006).

The rest of the paper is divided as follows. Section 2 presents our main results and the assumptions needed to derive them. Proofs are presented in the Appendix. Section 3 presents the results of a Monte-Carlo experiment. Section 4 presents an empirical analysis using the methods proposed here, confronting the performance of the bias-corrected average forecast with that of other types of forecast combination. Section 5 concludes.

2 Econometric Setup

Suppose that we are interested in forecasting a weakly stationary and ergodic univariate process $\{Y_t\}$ using a large number of forecasts that will be combined to yield an optimal forecast in the mean-squared error (MSE) sense. These forecasts could be the result of using several econometric models that need to be estimated prior to forecasting, or the result of using no formal econometric model at all, e.g., just the result of an opinion poll on the variable in question using a large number of individual responses.

We consider 3 consecutive distinct time periods, where time is indexed by $t = 1, 2, \ldots, T_1 \ldots, T_2 \ldots, T$. The period from $t = 1, 2, \ldots, T_1$ is labeled the "estimation sample," where models are usually fitted to forecast Y_t, if that is the case. The period from $t = T_1 + 1, \ldots, T_2$ is labeled the post-model-estimation or "training sample", where realizations of Y_t are usually confronted with forecasts produced in the estimation sample, if that is the case. The final period is $t = T_2 + 1, \ldots, T$, where genuine out-of-sample forecasting is entertained, benefiting from the results obtained during the training sample. In what follows, we let T_2 be $O(T)$. In order to guarantee that the number of observations in the training sample will go to infinity at rate T, we let T_1 be $O(1)$. Hence, asymptotic results will not hold for the estimation sample.

Regardless of whether forecasts are the result of a poll or of the estimation of econometric models, we label forecasts of Y_t, computed using conditioning sets lagged h periods, by $f_{i,t}^h$, $i = 1, 2, \ldots, N$. Therefore, $f_{i,t}^h$ are h-step-ahead forecasts and N is either the number of models estimated to forecast Y_t or the number of respondents of an opinion poll regarding Y_t.

In what follows we will let N go to infinity, which raises the question of whether this is plausible in our context. On the one hand, if forecasts are the result of estimating econometric models, they will differ across i if they are either based upon different conditioning sets or upon different functional forms of the conditioning set (or both). Since there is an infinite number of functional forms that could be entertained for forecasting, this gives an infinite number of possible forecasts. On the other hand, if forecasts are the result of a survey, although the number of responses is bounded from above, for all practical purposes, if a large enough number of responses is obtained, then the behavior of forecast combinations will be very close to the limiting behavior when $N \to \infty$.

We will focus on the following decomposition of Y_t:

$$Y_t = f_{i,t}^h - K_i - \varepsilon_{i,t}, \qquad i = 1, 2, \ldots, N, \qquad t > T_1, \tag{1}$$

where K_i is a time-invariant forecast bias of model i or of respondent i. This makes the error term $\varepsilon_{i,t}$ a zero-mean process, although it will be serially correlated in general. Because $f_{i,t}^h$ is an h-step-ahead forecast, K_i and $\varepsilon_{i,t}$ are respectively the h-step-ahead forecast bias and the forecast error associated with either model i or respondent i. Here, to simplify notation, we do not use an h superscript on K_i and $\varepsilon_{i,t}$, although they clearly depend on h.

At this point, it is desirable to discuss the nature of the term K_i. In particular, it is important to explain why we need to model it, which is related to the question of why we cannot focus solely on unbiased forecasts, for which $K_i = 0$. At first sight, (1) looks like an identity, but it is not, since we may also have a time-varying bias term. Therefore, the role of K_i in (1) is to capture the long-run effect in the time dimension of model-bias misspecification (econometric models of Y_t) or the long-run effect in the time dimension of the bias of respondent i. When considering econometric models, it is natural to assume that we do not know the data-generating process of Y_t. Therefore, all models that we might consider are inherently misspecified. In this case, K_i captures the long-run effect, in the time dimension, of forecast-bias misspecification of model i. By the same token, in the case of surveys, some of the surveyors may gain something by having a biased forecast. An interesting example in finance is that of a bank selling an investment fund. In this case, the bank's forecast of the fund return may be upward-biased simply because it may use this forecast as a marketing strategy to attract new clients; see Laster, Bennett and Geoum (1999). Patton and Timmermann (2006) consider an additional reason for the existence of K_i – the fact that there is uncertainty about the type of loss function used by forecasters in forming a specific forecast. There, forecasts that are unbiased under the loss function used by the forecaster may look biased under a different loss function.

Next, we have an assumption on how K_i relates to $\varepsilon_{i,t}$.

Assumption 1: We assume that $\mathbb{E}(\varepsilon_{i,t}|K_i) = 0$, for all t and that K_i is an identically distributed random variable in the cross-sectional dimension, but not necessarily independent[2], i.e.,

$$K_i \sim id(B, \sigma_k^2), \tag{2}$$

where B and σ_k^2 are respectively the mean and variance of K_i. It is important to distinguish between K_i and its realization k_i. In the time-series dimension, k_i has no variation, therefore, it is a fixed parameter.

The error term $\varepsilon_{i,t}$ is assumed to be weakly stationary and ergodic, reflecting the fact that, if forecasts are such that $\varepsilon_{i,t}$ is not weakly stationary and ergodic, then these forecasts could be simply discarded[3]. Because forecasts are computed h-steps ahead, forecast errors are serially correlated in general even if they are unbiased. Forecast errors are also likely to be

[2] The assumption of dependence is consistent with the idea that forecasters learn from one another by meeting, discussing, debating, and reading each other's analyses. Through their ongoing interactions, forecasters maintain a current, collective understanding of where the economy is most likely heading and its upside and downside risks.

[3] It is beyond the scope of this paper to discuss forecast combination for non-stationary processes. Also, note that although Y_t and $\varepsilon_{i,t}$ are ergodic for the mean, $f_{i,t}^h$ is non ergodic since K_i is a random variable.

cross-sectionally correlated, since the information set used by different models tends to overlap and poll responses tend to be similar for respondents with similar characteristics. In order to limit the degree of time-series and cross-sectional dependence of the errors, we assume the following:

Assumption 2: Let $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, \ldots \varepsilon_{N,t})'$ be an $N \times 1$ vector stacking the errors associated with all possible forecasts. Then, the vector process $\{\boldsymbol{\varepsilon}_t\}$ is assumed to be covariance-stationary and ergodic for the first and second moments, uniformly on N. Further, defining as $\xi_{i,t} = \varepsilon_{i,t} - \mathbb{E}_{t-1}(\varepsilon_{i,t})$, the innovation of $\varepsilon_{i,t}$, where $\mathbb{E}_{t-1}(\cdot)$ denotes the conditional expectation operator, we assume that

$$\lim_{N\to\infty} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| \mathbb{E}\left(\xi_{i,t}\xi_{j,t}\right) \right| = 0. \tag{3}$$

Assumption 2 controls the degree of time-series and cross-sectional dependence in the data. It does not rule out errors displaying conditional heteroskedasticity, since the latter can coexist with the assumption of weak stationarity; see Engle (1982). A similar assumption is made in Araujo, Issler and Fernandes (2006) to control the time-series and the cross-sectional decay within the framework of factor models applied to finance. Following the forecasting literature with large N and T, e.g., Stock and Watson (2002b), and the financial econometric literature, e.g., Chamberlain and Rothschild (1983), the condition $\lim_{N\to\infty} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| \mathbb{E}\left(\xi_{i,t}\xi_{j,t}\right) \right| = 0$ simply controls the degree of cross-sectional dependence present in forecast errors. It is noted by Bai (2005, p. 6), that Chamberlain and Rothschild's cross-sectional error decay requires:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| \mathbb{E}\left(\xi_{i,t}\xi_{j,t}\right) \right| < \infty. \tag{4}$$

Notice that this is the same cross-sectional decay used in Stock and Watson. Of course, (4) implies (3), but the converse is not true. Hence, Assumption

2 has a less restrictive condition than those commonly employed for factor models. It guarantees convergence in probability of cross-sectional means, which is why we use it here.

We start the discussion on forecast combination by solving (1) for $f_{i,t}^h$:

$$f_{i,t}^h = Y_t + K_i + \varepsilon_{i,t}, \qquad i = 1, 2, \ldots, N, \qquad t > T_1. \tag{5}$$

Equation (5) shows that we can decompose all forecasts into a common component Y_t, and two idiosyncratic components K_i and $\varepsilon_{i,t}$. The series being forecast (Y_t) is a common feature, in the sense of Engle and Kozicki(1993), of all forecasts. For any two series, a common feature exists if it is present in both of them and can be removed by linear combination. Here, subtracting any two forecasts eliminates Y_t. Araujo, Issler and Fernandes (2006) exploit this property to develop an estimator for the stochastic discount factor within a panel-data context. Here, we also exploit this property of Y_t in devising an optimal predictor for its realizations. We now state our first result.

Proposition 1 *If Assumptions 1 and 2 hold, then, the bias-corrected average forecast obeys* $\underset{N\to\infty}{plim}\left(\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h - \frac{1}{N}\sum_{i=1}^{N} k_i\right) = y_t$ *and* $\lim_{N\to\infty} MSE\left(\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h - \frac{1}{N}\sum_{i=1}^{N} k_i\right) = 0$, $t > T_1$, *where* $MSE(\cdot)$ *denotes the mean-squared error in forecasting* Y_t, y_t *denotes period-t realization of* Y_t, *and* k_i *denotes the realization of* K_i.

Proof. See Appendix. ■

Proposition 1 implies that the bias-corrected average forecast is an optimal forecast as N goes to infinity in the MSE sense. Because the MSE is bounded from below at zero, it has an MSE as small as that of any other individual forecast or as that of any other type of forecast combination. The latter is an important result which discourages estimating "optimal weights" in situations where N is large. This stems from the fact that, for large N, forecast combinations using "optimal population weights" and bias-corrected equal weights ($1/N$) will both have a zero MSE. However, in practice, one cannot resort to "optimal population weights," but rather has to estimate "optimal weights" from the data. Since the estimation period is fixed, $t = 1, 2, \ldots, T_1$ (although the training period is not), the performance of "optimal estimated weights" will not be as good as that of "optimal population weights," which explains the poor performance of "optimal estimated weights" compared with bias-corrected equal weights. From that perspective, there is no "forecast-combination puzzle."

Understanding the puzzle required using a weak law-of-large-numbers in a panel-data context. We see this as a major advantage of our approach *vis-à-vis* the commonly employed time-series approach with fixed N. Only in a panel-data framework can we formally state a weak law-of-large-numbers for forecast combinations and take full advantage of asymptotic results in both N and T. The lack of a broad of use of panel-data analysis in forecasting so far has limited our understanding of important phenomena in this literature. Of course, the lead of Stock and Watson(1999 and 2002a and b) and Forni et al. (2000, 2003) towards panel data has shed light on several important results on pooling information. We hope that our work will do the same as far as the pooling of forecasts is concerned.

One important feature of $\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^{h} - \frac{1}{N}\sum_{i=1}^{N} k_i$ is that it is unfeasible, since we do not observe the k_i's. Therefore, below we propose replacing k_i by a consistent estimator. The underlying idea behind the consistent estimator of k_i is that in the training sample one observes the realizations of Y_t and of the double-index process $f_{i,t}^{h}$, $i = 1...N$, and $T_1 < t < T_2$. Hence, one can form a panel of forecasts:

$$\left(f_{i,t}^{h} - y_t\right) = k_i + \varepsilon_{i,t}, \qquad i = 1, 2, \ldots, N, \qquad T_1 < t < T_2, \tag{6}$$

where i indexes forecasts, t indexes time, and it becomes obvious that k_i represents the fixed effect on this panel. It is natural to exploit this property of k_i in constructing a consistent estimator. This is exactly the approach taken here. In what follows, we propose a non-parametric estimator of k_i.

It does not depend on any distributional assumption on $K_i \sim id(B, \sigma_k^2)$ and it does not depend on any knowledge of the models used to compute the forecasts $f_{i,t}^h$. This feature of our approach widens its application to situations where the "underlying models are not known, as in a survey of forecasts" – Kang (1986); see also our empirical-application section.

Due to the nature of our problem – large number of forecasts – and the nature of k_i in (6) – time-invariant bias term – we need to consider large N, large T asymptotic theory to devise a consistent estimator for k_i. Panels with such a character are different from large N, small T panels. In order to allow the two indices N and T to pass to infinity jointly, we could consider a monotonic increasing function of the type $T = T(N)$, known as diagonal-asymptotic method; see Quah (1994) and Levin and Lin (1993). One drawback of this approach is that the limit theory that is obtained depends on the specific relationship considered in $T = T(N)$. A joint-limit theory allows both indices (N and T) to pass to infinity simultaneously without imposing any specific functional-form restriction. Despite that, it is substantially more difficult to derive and will usually apply only under stronger conditions, such as the existence of higher moments.

Searching for a method that allows robust asymptotic results without imposing too many restrictions (on functional relations and the existence of higher moments), we consider the sequential asymptotic approach developed by Phillips and Moon (1999). There, one first fixes N and then allows T to pass to infinity using an intermediate limit. Phillips and Moon write sequential limits of this type as $(T, N \to \infty)_{\text{seq}}$.

In order to clarify the idea behind sequential asymptotics, consider the following double-indexed process:

$$X_{N,T} = \frac{1}{k_N} \sum_{i=1}^{N} Z_{i,T},$$

and denote by Z_i the limit of $Z_{i,T}$ as $T \to \infty$. Phillips and Moon derive the sequential limit of $X_{N,T}$ as follows. By passing $T \to \infty$ for fixed N, an intermediate limit $X_N = \frac{1}{k_N}\sum\limits_{i=1}^{N} Z_i$ is found. Then, by letting $N \to \infty$ and by applying an appropriate limit theory to the standardized sum $X_N = \frac{1}{k_N}\sum\limits_{i=1}^{n} Z_i$, the final sequential limit is obtained. When $k_N = N$, this results in a law-of-large numbers being applied. When $k_N = \sqrt{N}$, this results in a central-limit theorem being applied.

In general, Phillips and Moon argue that a joint limit (N and T go to infinity simultaneously) is a more robust result than a sequential limit. However, these two could be equivalent. Following the intuition behind the convergence of a double-indexed real-number sequence, they show that if first-stage convergence in the sequential limit holds uniformly on the other index, then the sequential limit is equivalent to a joint limit, e.g., if $X_{N,T}$ converges to X_N uniformly in N, as $T \to \infty$, then sequential limit of $X_{N,T}$ is the same as the joint limit of $X_{N,T}$. Sequential panel-data asymptotics were applied in Phillips et al. (2001) and in Lima and Xiao (2007), among others. By using the sequential-limit approach, we can now state the second result of this paper.

Proposition 2 *If Assumptions 1 and 2 hold, the feasible bias-corrected average forecast obeys* $\underset{(T,N\to\infty)_{\text{seq}}}{plim}\left(\frac{1}{N}\sum\limits_{i=1}^{N} f_{i,t}^{h} - \frac{1}{N}\sum\limits_{i=1}^{N}\widehat{k}_i\right) = y_t$ *and* $\lim\limits_{(T,N\to\infty)_{\text{seq}}} MSE\left(\frac{1}{N}\sum\limits_{i=1}^{N} f_{i,t}^{h} - \frac{1}{N}\sum\limits_{i=1}^{N}\widehat{k}_i\right) = 0,\ t > T_1,$ *where* $\widehat{k}_i = \frac{1}{T}\sum_{t=1}^{T} f_{i,t}^{h} - \frac{1}{T}\sum_{t=1}^{T} y_t$ *is a consistent estimator of* k_i*, as* $T \to \infty$.

Proof. See Appendix. ■

In what follows we make explicit the role of $\widehat{B}$, the consistent estimator of B proposed in the proof of Proposition 2..

Proposition 3 *If Assumptions 1 and 2 hold, then,* $\underset{(T,N\to\infty)_{\text{seq}}}{plim}\left(\widehat{B} - B\right) = 0.$

Proof. See Appendix. ■

The results above provide important tools for large N, T forecasting. To get optimal forecasts, in the MSE sense, one has to combine all forecasts using simple averaging, appropriately centering it by using a bias-correction term. It is important to stress that, even though $N \to \infty$, the number of estimated parameters is kept at unity: $\widehat{B}$. This is a very attractive feature compared to models that combine forecasts estimating optimal weights. There the number of estimated parameters increases at the same rate as N, a clear disadvantage from the point of view of obtaining a small forecast variance.

The feasible bias-corrected average forecast can be made an even more parsimonious estimator of y_t when there is no need to estimate B. Of course, this raises the issue of whether $B = 0$, in which case the optimal forecast becomes $\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^{h}$ – the simple forecast combination originally proposed by Bates and Granger (1969). We next propose the following test statistic for $H_0 : B = 0$.

Proposition 4 *Under the null hypothesis $H_0 : B = 0$, the test statistic:*

$$\widehat{t} = \frac{\widehat{B}}{\sqrt{\widehat{V}}} \quad \xrightarrow[(T,N\to\infty)_{\text{seq}}]{d} \mathcal{N}(0,1),$$

where $\widehat{V}$ is a consistent estimator of the asymptotic variance of $\overline{B} = \frac{1}{N}\sum_{i=1}^{N} k_i$.

Proof. See Appendix. ■

3 Monte-Carlo Study

We considered the following data-generating process (DGP):

$$\begin{aligned} Y_t &= \beta X_t + \eta_t \\ t &= 1, 2, ..., T_1, ..., T_2, ...T, \end{aligned}$$

where $\{X_t\}$ is an $ARFIMA(1, d, 0)$ process, in which $(1 - L)^d X_t = \mu_t$, $(1 - \phi L)\mu_t = \epsilon_t$, and $\epsilon_t \sim iid\ N(0, 1)$. We also assumed that $\eta_t \sim iid\ N(0, 1)$ and η_t and ϵ_t are mutually independent. The data were generated by using functions of the ARFIMA package (Doornik and Ooms, 2001) for Ox programming language. In particular, we set $\phi = 0.8$, $\beta = 5$ and $d = 0.1$, 0.4, 0.49.

According to Beran (1998), one of the typical features of the above DGP (a stationary long-memory process) is that it generates local trends and cycles, but these are potentially spurious and disappear after some time. Therefore, for short sample sizes, we expect that this property will lead to poor forecasts since the estimated models will only capture the spurious local trend or cycle, which does not represent the true dynamics of Y_t. For this reason, we paid special attention to models that are estimated with a small sample. In this experiment, we considered the estimation sample to be as small as $T_1 = 30, 60, 120$, which are sample sizes commonly found in applied macroeconometrics.

As for the training sample, $T_2 - T_1$, this experiment included $T_2 - T_1 = 30, 60$. Recall that $T_2 = O(T)$ and, therefore we need to increase T to accommodate larger training samples. In this experiment, we set the number of out-of-sample observations as $(T - T_2) = 10$. Hence, when $T_1 = 30$ and $T_2 - T_1 = 30$, the total sample size T will be equal to 70. If $T_1 = 30$, but the training samples goes up to 60, then T must be equal to 100, etc.

We fitted the following auto-regressive distributed-lag models for Y_t,

$$\begin{aligned} Y_t &= c_0 + \sum_{j=1}^{J} \alpha_j Y_{t-j} + \sum_{i=0}^{I} \beta_i X_{t-i} + \epsilon_t \qquad (7) \\ for\ J &= 1, 2, ..., 6, \\ I &= 1, 2, ..., 5. \end{aligned}$$

For each J, we estimated a model with $I = 1, 2, ..., 5$, respectively. In all, we have estimated 30 models. We considered three forecast methods explained below, in addition to the simple average forecast (equal weights $1/N$ without any bias correction):

1. We estimated each model using observations up to period T_1, which is our estimation sample. The estimated models are next used to make one-step-ahead forecasts ($h = 1$) in the training sample, from $T_1 + 1$ to T_2. These forecasts are then used to estimate the model bias and the average bias. Without updating the estimation sample, each model is used to forecast observations from $T_2 + 1$ to T. Finally, the bias-corrected average forecasts is computed.

2. The same procedure as in 1 above is implemented without any bias correction: this is the simple average forecast combination.

3. After estimating all models in (7) using the estimation sample, the training sample is used to estimate the weights according to the following OLS regression:

$$y_s = \omega_0 + \sum_{i=1}^{30} \omega_{it} f_{i,s}^1 + \varepsilon_s,\ s = T_1 + 1, ...T_2, \tag{8}$$

where $f_{i,s}^1$ is the one-step ahead forecast made by the i-th model for the s-th observation in the training sample. The estimated weights are used to compute the weighted-average forecast from $T_2 + 1$ to T.

We then compute the MSE of the simple average forecast, the bias-corrected average forecast, and the weighted-average forecast for the periods $T_2 + 1, ..., T$. All the forecasts are one-step-ahead static forecasts, i.e., forecasts for $t + 1$ used observed data for Y_t.

The number of Monte-Carlo replications was set to $5,000$. For each replication, 30 models are estimated and forecasts are made according to each of the three aforementioned strategies. After $5,000$ replications, we computed two distributions of relative MSEs, using the bias-corrected average forecast as *numeraire*. The first distribution is that of the relative MSE of the average forecast (the MSE of the simple average forecast divided by that of the bias-corrected average forecast). The second is that of the weighted-average forecast (the MSE of the weighted-average forecast divided by that of the bias-corrected average forecast).

We report the mean of each distribution in Table 1. The notation $RMSE_i^p$ $i = 1, 2$ denotes the mean of the relative MSE of the average forecast ($RMSE_1^p$) and weighted-average forecast ($RMSE_2^p$). The superscript p indicates the number of observations in the training sample.

The results in Table 1 show that, for estimation sample as small as $T_1 = 30$, the bias-corrected average forecast outperforms the simple average forecast. In particular, such advantage increases as the presence of long memory is stronger, that is, as the fractional-integration parameter d increases. Indeed, for $d = 0.1$, $RMSE_1^{30} = 1.23$, whereas $RMSE_1^{30} = 1.65$ when $d = 0.49$.[4] The forecast can be improved if more observations are used in the training sample. For example, when 120 observations are used to compute the average bias, we obtain $RMSE_1^{120} = 1.75$ for $d = 0.49$.

The good performance of the bias-corrected average forecast results from the fact that stationary long-memory processes generate local trends and cycles that disappear only after a long time. For short-estimation samples, the econometric models (7) will all include irrelevant regressors, which may lead to non-trivial forecast biases, the smaller the estimation sample. Of course, as the estimation sample increases, the coefficients of these irrelevant regressors will be approximately zero and we should expect the gains of bias correction to decrease. Our Monte-Carlo experiment shows that the method proposed in this paper can improve forecast accuracy by estimating and removing this short-sample forecast bias. As the estimation sample increases, say, to $T_1 = 120$, the local trends and cycles become less important and therefore the misspecification problem diminishes.[5] As a result, the econo-

[4] Recall that a long-memory process is stationary as long as $0 < d < 0.5$.

[5] As motivated by Hendry and Clements (2002), model bias can be corrected by including intercept in the regressions. Notice, however, that when the estimation sample is short, the spurious trend or cycle (generated by the presence of long memory) becomes influential, and the estimation of the intercept may also be strongly biased.

metric models will give rise to small bias and, consequently, bias-correction will not be as important as in the case of a small estimation sample.

A striking result presented in Table 1 is the effect of long memory on the performance of the weighted-average forecast. Such forecast method has the worst performance among the three methods considered. Its performance deteriorates significantly when the fractional-integration parameter increases. This result suggests that local trends and cycles generated by a stationary long-memory process (along with a small estimation sample) strongly bias the estimation of the "optimal" weights used to compute the weighted-average forecast. In this way, the well-known forecast-combination puzzle may simply be a reflection of the potential misspecification of econometric models used in forecasting.

4 Empirical Application

Professional forecasts guide market participants and inform them about future economic conditions. However, many analysts argue that forecasters might strategically bias forecasts as long as they receive economic incentives to do so. The importance of microeconomic incentives for forecasters and analysts is stressed by a number of empirical studies, such as Ehrbeck and Waldmann (1996), Graham (1999), Hong et al. (2000), Lamont (2002), Welch (2000), and Zitzewitz (2001).

In this section, we present an application of the method proposed here for the case of forecast surveys, focusing on two different surveys. Our results show that bias correction can indeed help forecasting. We also test the hypothesis that professional forecasters behave strategically in a statistical sense, perhaps because they earn more from forecast publicity than from forecast accuracy. When this is accounted for, the bias-corrected average forecast introduced in this paper outperforms simple forecast averages (consensus). It is important to stress that, although our techniques were conceived for a large N, T environment, the empirical results presented here show the usefulness of our method even in a small N, T environment. Also, the forecasting gains from bias correction, whenever the average forecast is biased, are non-trivial.

4.1 Philadelphia Fed's Survey of Professional Forecasters

In our first empirical application, we consider a panel data of individual responses from the Philadelphia Fed's Survey of Professional Forecasters available at quarterly frequency. The forecasters in this Survey come largely from the business world and Wall Street. One important feature of this Survey is anonymity of the institution supplying a given forecast. This is designed to encourage forecasters to provide their best forecast without fearing the consequences of making mistakes.

For a long time it has been common knowledge that the average forecast usually performs better than alternative forecast combinations when survey data is used. In fact, Kang (1986) concludes that "A simple average should be used when underlying models are not known, as in a survey of forecasts..." In this section we show how a macroeconomist using survey data can use the bias-corrected average forecast to improve upon the consensus (average) forecast.

In order to construct our panel of forecasts, we have to consider the fact that many forecasters report missing values for different reasons, which is a problem in trying to obtain a long-balanced panel of forecasts. To that end, we included forecasters who reported nine consecutive one-step-ahead forecasts from 2002:4 to 2004:4. Hence, our time dimension is $T = 9$. To compute the average bias, we compared one-step-ahead forecasts with realizations of the forecasting variables from 2002:4 through 2003:4, comprising 5 observations. Therefore, we were left with observations from 2004:1 through 2004:4 for out-of-sample forecast evaluation (4 observations).

We focused our exercise on two important macroeconomic variables covered by this survey: the annualized CPI inflation rate and the unemployment rate, both seasonally adjusted. For CPI inflation we observed a maximum of $N = 19$ forecasters, whereas for the unemployment rate, we observed a maximum of $N = 22$ forecasters.

Table 2 exhibits the estimate of the sample average bias for CPI inflation and the Unemployment Rate. We also test whether the average bias is zero (p-values in parenthesis). Our estimates reveal a negative average bias for the annualized inflation rate and a positive bias for the unemployment rate. Both are statistically significant at the 10% level, but only the unemployment average bias is significant at the 5% level. This result suggests that we could use these average-bias estimates to improve forecasting, although we should expect a larger improvement in the case of unemployment. Indeed, the simple average forecast is 9% worse than the feasible bias-corrected average forecast for CPI inflation and 56% worse for the Unemployment Rate.

4.2 The Central Bank of Brazil's "Focus Forecast Survey"

The "Focus Forecast Survey," organized by the Central Bank of Brazil, is a unique panel database of forecasts. It collects forecast information on almost 120 institutions, including commercial banks, asset managers and non-financial institutions, which are followed throughout time. Forecasts have been collected since 1998, which potentially can serve to approximate a large N, T environment. Besides that, it also has the following desirable features: the anonymity of forecasters is preserved, although the names of the top-five forecasters for a given economic variable is released by the Central Bank of Brazil; forecasts are collected at different frequencies (monthly, semi-annual, annual), as well as at different forecast horizons (e.g., short-run forecasts are obtained for h from 1 to 12 months); there is a large array of macroeconomic time series included in the survey.

To save space, below we focus our analysis only on the behavior of forecasts of the monthly inflation rate (π_t) in Brazil, measured by the official

Consumer Price Index (CPI), computed by FIBGE. In order to obtain the largest possible balanced panel ($N \times T$), we used $N = 18$ and a time-series sample period covering the period 2002:11 through 2006:3 ($T = 41$). Of course, in the case of a survey panel, there is no estimation sample. We chose the first 23 time observations to compute $\widehat{B}$ – the average bias – leaving 18 time-series observations for out-of-sample forecast evaluation. The forecast horizon chosen was $h = 6$, this being an important horizon to determine future monetary policy within the Brazilian Inflation-Targeting program.

The results of our empirical exercise are presented in Tables 3 and 4. They show that the average bias is positive for the 6-month horizon – about 0.075 – and significant at the 10% level, with a p-value of 0.09. Out-of-sample forecast comparisons between the simple average and the bias-corrected average forecast show that the former has an MSE 11% bigger than that of the latter.

5 Conclusions and Extensions

In this paper, we propose a novel approach to econometric forecasting of stationary and ergodic series within a panel-data framework, where the number of forecasts and the number of time periods increase without bounds. The advantages of our approach are many. First, only in an asymptotic panel-data context we can fully understand why the pooling of forecasts works in practice. Second, we can also propose improvements on simple forecast-combination schemes – such as the simple forecast combination. Here, we propose the bias-corrected average forecast. Third, our techniques are applicable in two important contexts: when forecasts are a result of model estimation, and when they are the result of opinion polls. Fourth, the method proposed here is non-parameteric: it requires no distributional assumption whatsoever on the variables involved, and also no knowledge of the models used in forecasting.

The basis of our technique is to decompose individual forecasts into three components: the series being forecast, a time-invariant forecast bias, and a zero-mean forecast error. The series being forecast is viewed as a *common feature* of all individual forecasts. Standard tools from panel-data asymptotic theory are then used to devise an optimal forecasting combination that has a zero limiting mean-squared forecast error. This optimal forecast combination uses equal weights and a bias-correction term. The use of equal weights avoids estimating forecast weights, which contributes to reduce forecast variance, although potentially at the cost of an increase in bias. The use of a bias-correction term eliminates any possible detrimental effect arising from equal weighting. We label this optimal forecast as the *bias-corrected average forecast*.

In theory – large N and T – the use of a bias-corrected average forecast is potentially superior to the use of any single forecast and is equal or superior to any other combining technique. Moreover, in practice – small N and/or T – an important element of the use of the bias-corrected average forecast is that the forecast combination puzzle works to our advantage, now augmented with a bias-correction term. Hence, there will be situations in which we can improve upon the simple average forecast by using a bias-correction, and others which we cannot. Our framework offers a statistical test for excluding the bias-correction term.

The Monte-Carlo experiment and the empirical analyses performed here show the usefulness of our new approach. Regarding model misspecification bias, the Monte-Carlo experiment shows important improvement over conventional combination techniques – from about 5% to about 75%. compared to the simple forecast combination under MSE loss. A much larger improvement is observed in the case of "optimal forecasting weights." In the empirical exercise, we showed that using our method leads to an improvement in forecasting accuracy under MSE loss – from about 10% to about 60% relative to the simple forecast combination under MSE loss. As one should expect, higher gains for bias correction are observed when the null hypothesis of a zero bias is rejected in testing.

For reasons of space, we refrain from fully discussing here natural extensions of our proposed method. A partial account of those includes the following:

1. In the panel of forecasts:

$$\left(f_{i,t}^{h} - y_t\right) = k_i + \varepsilon_{i,t}, \qquad i = 1, 2, \ldots, N, \qquad T_1 < t < T_2, \tag{9}$$

   we imposed a unity coefficient for y_t, but we could have had an encompassing panel-regression system:

$$f_{i,t}^{h} = \beta_i y_t + k_i + \varepsilon_{i,t}, \qquad i = 1, 2, \ldots, N, \qquad T_1 < t < T_2, \tag{10}$$

   where β_i can be interpreted as the *beta* of forecast-model i vis-à-vis y_t. A natural hypothesis to test is $H_0 : \beta_i = 1$, for all i, which can be implemented using standard panel techniques.

2. There may be instances where forecast models produce forecasts that are too highly correlated. In theory, this may prevent a weak law-of-large-numbers from holding for the error terms. In this case we can combine pooling of information and pooling of forecasts:

$$\left(f_{i,t}^{h} - y_t\right) = k_i + \sum_{k=1}^{K} \beta_{i,k} f_{k,t} + \eta_{i,t}, \qquad i = 1, 2, \ldots, N, \qquad T_1 < t < T_2, \tag{11}$$

   where $f_{k,t}$ are zero-mean pervasive factors and, as is usual in factor analysis,
   $\underset{N \to \infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} \eta_{i,t} = 0$. In this context, we implemented the following decomposition:

$$\varepsilon_{i,t} = \sum_{k=1}^{K} \beta_{i,k} f_{k,t} + \eta_{i,t}, \qquad i = 1, 2, \ldots, N, \qquad T_1 < t < T_2.$$

   Thus, factor and principal-component analyses (Stock and Watson(1999 and 2002a and b) and Forni et al. (2000, 2003)) are combined with the idea of bias-corrected average forecasts. Hence, we could combine *pooling of forecasts* with *pooling of information* within the same model.

3. The final extension considered here is to allow for a time-varying bias term ξ_t. In this case,

$$\left(f_{i,t}^h - y_t\right) = k_i + \xi_t + \varepsilon_{i,t}, \qquad i = 1, 2, \ldots, N, \qquad T_1 < t < T_2. \quad (12)$$

The techniques of Fuller and Battese (1974) can be a starting point to generate consistent estimates of k_i and ξ_t in a context where N and T are large.

References

[1] Araujo, F., Issler, J.V. and Fernandes, M.(2006), "A Stochastic Discount Factor Approach to Asset Pricing Using Panel Data," Working Paper: EPGE-FGV # 628, downloadable from http://epge.fgv.br/portal/arquivo/2155.pdf

[2] Bai, J., (2005), "Panel Data Models with Interactive Fixed Effects," Working Paper: New York University.

[3] Bai, J., and S. Ng, (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191-221.

[4] Bai, J. and S. Ng, (2004), "Evaluating Latent and Observed Factors in Macroeconomics and Finance," Working Paper: University of Michigan.

[5] Beran, Jan, (1998), "*Statistics for Long-Memory Processes.*" Chapman and Hall – CRC.

[6] Chamberlain, Gary, and Rothschild, Michael, (1983). "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, vol. 51(5), pp. 1281-1304.

[7] Clements, M.P. and D.F. Hendry, 1998, *Forecasting Economic Time Series*, Cambridge University Press.

[8] Clements, M.P. and D.F. Hendry, 2006, Forecasting with Breaks in Data Processes, in C.W.J. Granger, G. Elliott and A. Timmermann (eds.) *Handbook of Economic Forecasting*, pp. 605-657, Amsterdam, North-Holland.

[9] Conley, T.G., 1999, "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, Vol. 92 Issue 1, pp. 1-45.

[10] Connor, G., and R. Korajzcyk (1986), "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis," *Journal of Financial Economics*, 15, 373-394.

[11] Connor, G. and Korajczyk, R. (1993), "A test for the number of factors in am approximate factor structure," *Journal of Finance* 48, 1263 - 1291.

[12] Doornik, J. and M. Ooms (2001), "*A package for estimating, forecasting and simulating arfima models: Arfima package 1.01 for ox.*" OX open source software..

[13] Ehrbeck, T., Waldman, R., 1996. Why are professional forecasters biased? Agency versus behavioral explanations. *Quarterly Journal of Economics* 111, 21-40.

[14] Elliott, G., C.W.J. Granger, and A. Timmermann, 2006, Editors, *Handbook of Economic Forecasting*, Amsterdam: North-Holland.

[15] Elliott, G. and A. Timmermann (2005), "Optimal forecast combination weights under regime switching", *International Economic Review*, 46(4), 1081-1102.

[16] Elliott, G. and A. Timmermann (2004), "Optimal forecast combinations under general loss functions and forecast error distributions", *Journal of Econometrics* 122:47-79.

[17] Engle, R.F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, pp. 987-1006.

[18] Engle, R. F., Issler, J. V., 1995, "Estimating common sectoral cycles," *Journal of Monetary Economics*, vol. 35, 83–113.

[19] Engle, R.F. and Kozicki, S. (1993). "Testing for Common Features", *Journal of Business and Economic Statistics*, 11(4): 369-80.

[20] Forni, M., Hallim, M., Lippi, M. and Reichlin, L. (2000), "The Generalized Dynamic Factor Model: Identification and Estimation", *Review of Economics and Statistics*, 2000, vol. 82, issue 4, pp. 540-554.

[21] Forni M., Hallim M., Lippi M. and Reichlin L., 2003 "The Generalized Dynamic Factor Model one-sided estimation and forecasting," *Journal of the American Statistical Association*, forthcoming.

[22] Forni, M., M. Hallim, M. Lippi and L. Reichlin (2004), "The generalized factor model: consistency and rates", *Journal of Econometrics* 119:231-255.

[23] Fuller, Wayne A. and George E. Battese, 1974, "Estimation of linear models with crossed-error structure," *Journal of Econometrics*, Vol. 2(1), pp. 67-78.

[24] Giacomini, Raffaella and Halbert White, 2006, "Tests of Conditional Predictive Ability," *Econometrica*, vol. 74(6), pp. 1545-1578.

[25] Graham, J., 1999. Herding among investment newsletters: theory and evidence. Journal of Finance 54, 231-268.

[26] Granger, C.W.J., and R.Ramanathan (1984), "Improved methods of combining forecasting", *Journal of Forecasting* 3:197–204.

[27] Hendry, D.F. and M.P. Clements (2002), “Pooling of forecasts”, *Econometrics Journal*, 5:1-26.

[28] Hong, H., Kubik, J.D., Solomon, A., 2000. Security analysts ´career concerns and herding of earning forecasts. *RAND Journal of Economics* 31, 121-144.

[29] Hoogstrate, Andre J., Franz C. Palm, Gerard A. Pfann, 2000, “Pooling in Dynamic Panel-Data Models: An Application to Forecasting GDP Growth Rates,” *Journal of Business and Economic Statistics*, Vol. 18, No. 3, pp. 274-283.

[30] Issler, J. V., Vahid, F., 2001, “Common cycles and the importance of transitory shocks to macroeconomic aggregates,” *Journal of Monetary Economics*, vol. 47, 449–475.

[31] Issler, J. V., Vahid, F., 2006, “The missing link: Using the NBER recession indicator to construct coincident and leading indices of economic activity,” *Annals Issue of the Journal of Econometrics* on *Common Features*, vol. 132(1), pp. 281-303.

[32] Kang, H. (1986), "Unstable Weights in the Combination of Forecasts," *Management Science* 32, 683-95.

[33] Lamont, O., 2002. Macroeconomic forecasts and microeconomic forecasters. *Quarterly Journal of Economics* 114, 293-318.

[34] Levin, A., Lin, C.-F., 1993, “Unit root tests in panel data:asymptotic and finite-sample properties,” UC San Diego: Unpublished Working Paper.

[35] Lima, L.R. and Z. Xiao, 2007. "Do Shocks Last Forever? Local Persistency in Economic Time Series". *Journal of Macroeconomics, vol.29, n. 2.*

[36] Pesaran, M.H., (2005), "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure." Working Paper: Cambridge University, forthcoming in *Econometrica*.

[37] Phillips, P.C.B., 1995, "Lecture Notes," downloadable from the author's webpage: Yale University.

[38] Phillips, P.C.B. and H.R. Moon, 1999, "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, vol. 67 (5), pp. 1057–1111.

[39] Phillips, P.C.B., H.R. Moon and Z. Xiao, 2001, "How to estimate autoregressive roots near unity," *Econometric Theory* 17, 29-69.

[40] Quah, D., 1994, "Exploiting cross section variation for unit root inference in dynamic data," *Economic Letters* 44, 9–19.

[41] Ross, S.A. (1976), "The arbitrage theory of capital asset pricing", *Journal of Economic Theory*, 13, pp. 341-360.

[42] Smith, Jeremy and Kenneth F. Wallis, 2005, "Combining Point Forecasts: The Simple Average Rules, OK?" Working Paper: Department of Economics, University of Warwick.

[43] Stock, J. and Watson, M., "Forecasting Inflation", *Journal of Monetary Economics*, 1999, Vol. 44, no. 2.

[44] Stock, J. and Watson, M., "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business and Economic Statistics*, April 2002a, Vol. 20 No. 2, 147-162.

[45] Stock, J. and Watson, M., "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 2002b.

[46] Stock, J. and Watson, M., 2006, "Forecasting with Many Predictors,"In: Elliott, G., C.W.J. Granger, and A. Timmermann, 2006, Editors, *Handbook of Economic Forecasting*, Amsterdam: North-Holland, Chapter 10, pp. 515-554.

[47] Timmermann, A., 2006, "Forecast Combinations," In: Elliott, G., C.W.J. Granger, and A. Timmermann, 2006, Editors, *Handbook of Economic Forecasting*, Amsterdam: North-Holland, Chapter 4, pp. 135-196.

[48] Vahid, F. and Engle, R. F., 1993, "Common trends and common cycles," *Journal of Applied Econometrics*, vol. 8, 341–360.

[49] Vahid, F., Engle, R. F., 1997, "Codependent cycles," *Journal of Econometrics*, vol. 80, 199–221.

[50] Vahid, F., Issler, J. V., 2002, "The importance of common cyclical features in VAR analysis: A Monte Carlo study," *Journal of Econometrics*, 109, 341–363.

[51] Welch, I., 2000. Herding among security analysts. *Journal of Financial Economics* 58, 369-396.

[52] Zitzewitz, E., 2001. Measuring herding and exaggeration by equity analysts. Unpublished working paper. Graduate School of Business, Stanford University.

A Proofs of Propositions in Section 2

Proof of Proposition 1. Because $\varepsilon_{i,t}$ is weakly stationary and mean-zero, for every i, there exists a scalar Wold representation of the form:

$$\varepsilon_{i,t} = \sum_{j=0}^{\infty} b_{i,j} \xi_{i,t-j} \tag{13}$$

where, for all i, $b_{i,0} = 1$, $\mu_i < \infty$, $\sum_{j=0}^{\infty} b_{i,j}^2 < \infty$, and $\xi_{i,t}$ is white noise. We consider now the sample cross-sectional average of equation (5):

$$\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h = y_t + \frac{1}{N}\sum_{i=1}^{N} k_i + \frac{1}{N}\sum_{i=1}^{N} \varepsilon_{i,t}, \tag{14}$$

and examine the convergence in probability ($N \to \infty$) of each term in (14). Under Assumption 1,

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} k_i = B.$$

We now examine the convergence in probability of $\frac{1}{N}\sum\limits_{i=1}^{N} \varepsilon_{i,t}$. Our strategy is to show that, in the limit, the variance of $\frac{1}{N}\sum\limits_{i=1}^{N} \varepsilon_{i,t}$ is zero, a sufficient condition for a weak law-of-large-numbers (WLLN) to hold for $\{\varepsilon_{i,t}\}$. In computing the variance of $\frac{1}{N}\sum\limits_{i=1}^{N}\sum\limits_{j=0}^{\infty} b_{i,j}\xi_{i,t-j}$ we use the fact that there is no cross correlation between $\xi_{i,t}$ and $\xi_{i,t-k}$, $k = 1, 2, \ldots$. Therefore, we need only to consider the sum of the variances of terms of the form $\frac{1}{N}\sum_{i=1}^{N} b_{ik}\xi_{i.t-k}$. These variances are given by:

$$\mathrm{VAR}\left(\frac{1}{N}\sum_{i=1}^{N} b_{i,k}\xi_{i,t-k}\right) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=0}^{N} b_{i,k}b_{j,k}\mathbb{E}\left(\xi_{i,t}\xi_{j,t}\right), \tag{15}$$

due to weak stationarity of $\boldsymbol{\varepsilon}_t$. We now examine the limit of the generic term in (15) with detail:

$$\mathrm{VAR}\left(\frac{1}{N}\sum_{i=1}^{N} b_{i,k}\xi_{i,t-k}\right) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} b_{i,k}b_{j,k}\mathbb{E}\left(\xi_{i,t}\xi_{j,t}\right) \leq$$

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \left|b_{i,k}b_{j,k}\mathbb{E}\left(\xi_{i,t}\xi_{j,t}\right)\right| = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \left|b_{i,k}b_{j,k}\right| \left|\mathbb{E}\left(\xi_{i,t}\xi_{j,t}\right)\right| \leq \tag{16}$$

$$\left(\max_{i,j}\left|b_{i,k}b_{j,k}\right|\right)\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \left|\mathbb{E}\left(\xi_{i,t}\xi_{j,t}\right)\right|. \tag{17}$$

Hence:

$$\lim_{N\to\infty} \text{VAR}\left(\frac{1}{N}\sum_{i=1}^{N} b_{i,k}\xi_{i,t-k}\right) \leq \lim_{N\to\infty}\left(\max_{i,j}|b_{i,k}b_{j,k}|\right) \times$$
$$\lim_{N\to\infty}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\mathbb{E}\left(\xi_{i,t}\xi_{j,t}\right)\right| = 0,$$

since the sequence $\{b_{i,j}\}_{j=0}^{\infty}$ is square-summable, yielding $\lim_{N\to\infty}\left(\max_{i,j}|b_{i,k}b_{j,k}|\right) \leq \infty$, and Assumption 2 imposes $\lim_{N\to\infty}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\mathbb{E}\left(\xi_{i,t}\xi_{j,t}\right)\right| = 0$.

Thus, all variances are zero in the limit, as well as their sum, which gives:

$$\begin{aligned} \underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t} &= 0, \text{ and,} \\ \underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^{h} - \frac{1}{N}\sum_{i=1}^{N} K_i\right) &= y_t, \end{aligned} \tag{18}$$

where y_t is the realization of Y_t. We are now ready to compute MSEs for $\frac{1}{N}\sum\limits_{i=1}^{N} f_{i,t}^{h} - \frac{1}{N}\sum\limits_{i=1}^{N} k_i$. In doing so, we confront realizations $\{y_t\}_{t=T_1+1}^{T}$ with $\left\{\frac{1}{N}\sum\limits_{i=1}^{N} f_{i,t}^{h} - \frac{1}{N}\sum\limits_{i=1}^{N} k_i\right\}_{t=T_1+1}^{T}$. However, as $N \to \infty$, these two sequences become identical. Therefore,

$$\lim_{N\to\infty} \text{MSE}\left(\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^{h} - \frac{1}{N}\sum_{i=1}^{N} k_i\right) = 0.$$

■

Proof of Proposition 2. We search for a feasible consistent estimator of the bias-corrected average forecast. This entails a feasible estimator for B, the mean of K_i. Although Y_t and $\varepsilon_{i,t}$ are ergodic for the mean, $f_{i,t}^{h}$ is non ergodic since $K_i \sim id(B, \sigma_k^2)$. This implies that:

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} f_{i,t}^{h} &= \frac{1}{T}\sum_{t=1}^{T} y_t + \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{i,t} + k_i \\
&\xrightarrow{p} \mathbb{E}\left(Y_t\right) + k_i + \mathbb{E}\left(\varepsilon_{i,t}\right) \\
&= \mathbb{E}(f_{i,t}^{h}|\Im),
\end{aligned}
$$

where $\Im$ is the invariant field spanned by K_i, (see Phillips 1995, for a more complete discussion). The last line makes clear the dependence of simple average forecast on the realizations of K_i, which explains why $f_{i,t}^{h}$ is non ergodic although Y_t and $\varepsilon_{i,t}$ are.

Using the fact that,

$$
\mathbb{E}\left(\varepsilon_{i,t}\right) = 0, \qquad \text{for } i = 1, 2, ..., N,
$$

we obtain:

$$
k_i = \mathbb{E}(f_{i,t}^{h}|\Im) - \mathbb{E}\left(Y_t\right).
$$

This leads us to propose the following consistent estimator for k_i,

$$
\begin{aligned}
\widehat{k}_i &= \frac{1}{T}\sum_{t=1}^{T} f_{i,t}^{h} - \frac{1}{T}\sum_{t=1}^{T} y_t, \quad i = 1, ..., N \\
&= \frac{1}{T}\sum_{t=1}^{T}(y_t + k_i + \varepsilon_{i,t}) - \frac{1}{T}\sum_{t=1}^{T} y_t \\
&= k_i + \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{i,t}. \text{ or,} \\
\widehat{k}_i - k_i &= \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{i,t}.
\end{aligned}
$$

Since $\frac{1}{T}\sum_{t=1}^{T} \varepsilon_{i,t} \xrightarrow{p} \mathbb{E}\left(\varepsilon_{i,t}\right) = 0$, we have that $\widehat{k}_i \xrightarrow{p} k_i$.

Notice that:

$$
\widehat{B} = \frac{1}{N}\sum_{i=1}^{N} \widehat{k}_i = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T} f_{i,t}^{h} - \frac{1}{T}\sum_{t=1}^{T} y_t\right]. \tag{19}
$$

Now, we can write the feasible bias-corrected average forecast as:

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^{h}-\widehat{B} &= \frac{1}{N}\sum_{i=1}^{N} f_{i,t}^{h}-\frac{1}{N}\sum_{i=1}^{N}\widehat{k}_i=\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^{h}-\frac{1}{N}\sum_{i=1}^{N}\left[k_i+\frac{1}{T}\textstyle\sum_{t=1}^{T}\varepsilon_{i,t}\right] \\ &= \frac{1}{N}\sum_{i=1}^{N} f_{i,t}^{h}-\frac{1}{N}\sum_{i=1}^{N}k_i+\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\textstyle\sum_{t=1}^{T}\varepsilon_{i,t}.\end{aligned}$$

By the argument of sequential asymptotics in Phillips and Moon (1999), we let $T\to\infty$ first. Since $\varepsilon_{i,t}$ is ergodic for the mean, $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{i,t}\xrightarrow{p}0$. However, in this case, as $N\to\infty$, the asymptotic behavior of $\frac{1}{N}\sum\limits_{i=1}^{N} f_{i,t}^{h}-\frac{1}{N}\sum\limits_{i=1}^{N}\widehat{k}_i$ will be identical of that of $\frac{1}{N}\sum\limits_{i=1}^{N} f_{i,t}^{h}-\frac{1}{N}\sum\limits_{i=1}^{N}k_i$. Now letting $N\to\infty$, we are back to the result in Proposition 1, proving Proposition 2.

As a final issue in the proof, it is worthwhile analyzing the last term in :

$$\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^{h}-\widehat{B}=\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^{h}-\frac{1}{N}\sum_{i=1}^{N}k_i+\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\textstyle\sum_{t=1}^{T}\varepsilon_{i,t}.$$

By passing $T\to\infty$ for fixed N, we note that intermediate limit holds uniformly on N, since the process $\{\varepsilon_{i,t}\}_{t=1}^{T}$ is ergodic for the mean, uniformly in N. Therefore, the sequential limit is equivalent to the joint limit, showing that we do not need to impose stringent moment restrictions on $\varepsilon_{i,t}$ to prove Proposition 2. ■

Proof of Proposition 3. Define, $\overline{B}=\frac{1}{N}\sum\limits_{i=1}^{N}k_i$. Then,

$$\left(\widehat{B}-\overline{B}\right)=\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{k}_i-k_i\right).$$

By the WLLN, as $T\to\infty$,

$$\begin{aligned}&\widehat{k}_i\xrightarrow{p}k_i,\text{ and,}\\ &\frac{1}{N}\sum_{i=1}^{N}\widehat{k}_i\xrightarrow{p}\frac{1}{N}\sum_{i=1}^{N}k_i.\end{aligned}$$

As $N \to \infty$,

$$\frac{1}{N}\sum_{i=1}^{N} k_i \xrightarrow{p} B,$$

where B is the mean of k_i of the cross-sectional distribution of k_i under Assumption 1.

Hence, as $(T, N \to \infty)_{\text{seq}}$,

$$\widehat{B} \xrightarrow{p} B, \text{ and } \overline{B} \xrightarrow{p} B \text{ as well. Then,}$$

$$\operatorname*{plim}_{(T,N\to\infty)_{\text{seq}}} \left(\widehat{B} - B\right) = 0.$$

■

Proof of Proposition 4. Under $H_0 : B = 0$, we have shown in Proposition 3 that $\widehat{B}$ is a $(T, N \to \infty)_{\text{seq}}$ consistent estimator for B. To compute the consistent estimator of the asymptotic variance of $\overline{B}$ we follow Conley(1999), who matches spatial dependence to a metric of *economic distance*. Denote by $\text{MSE}_i(\cdot)$ and $\text{MSE}_j(\cdot)$ the MSE in forecasting of forecasts i and j respectively. For any two generic forecasts i and j, we use $\text{MSE}_i(\cdot) - \text{MSE}_j(\cdot)$ as a measure of distance between these two forecasts. For N forecasts, we can choose one of them to be the benchmark, say, the first one, computing $\text{MSE}_i(\cdot) - \text{MSE}_1(\cdot)$ for $i = 2, 3, \cdots, N$. With this measure of spatial dependence at hand, we can construct a two-dimensional estimator of the asymptotic variance of $\overline{B}$ and $\widehat{B}$ following Conley(1999, Sections 3 and 4). We label $\overline{V}$ and $\widehat{V}$ the estimates of the asymptotic variances of $\overline{B}$ and of $\widehat{B}$, respectively.

Once we have estimated the asymptotic covariance of $\overline{B}$, we can test the null hypothesis $H_0 : B = 0$, by using the following t-ratio statistic:

$$t = \frac{\overline{B}}{\sqrt{\overline{V}}}.$$

By the central limit theorem, $t \xrightarrow[N\to\infty]{d} \mathcal{N}(0,1)$ under $H_0 : B = 0$. Now consider $\widehat{t} = \frac{\widehat{B}}{\sqrt{\widehat{V}}}$, where $\widehat{V}$ is computed using $\widehat{k} = (\widehat{k}_1, \widehat{k}_2, ..., \widehat{k}_N)'$ in place of $k = (k_1, k_2, ..., k_N)'$. We have proved that $\widehat{k}_i \xrightarrow{p} k_i$ as $T \to \infty$, then the test statistics t and $\widehat{t}$ are asymptotically equivalent and therefore

$$\widehat{t} = \frac{\widehat{B}}{\sqrt{\widehat{V}}} \quad \xrightarrow[(T,N\to\infty)_{\mathrm{seq}}]{d} \mathcal{N}\left(0,1\right).$$

■

B Tables and Figures

Table 1: Monte-Carlo Results

d	$RMSE_1^{30}$	$RMSE_2^{30}$	$RMSE_1^{120}$	$RMSE_2^{120}$
		$T_1 = 30$		
0.1	1.23	37.8	1.23	20.33
0.4	1.53	319	1.56	629.11
0.49	1.65	1,405	1.75	2,148
		$T_1 = 60$		
0.1	1.06	46.1	1.08	17.9
0.4	1.15	762	1.20	664
0.49	1.20	1,813	1.26	2,145
		$T_1 = 120$		
0.1	1.00	81.8	1.03	20.86
0.4	1.05	565.1	1.07	1,970
0.49	1.07	8,942	1.11	1,235

Notes: (A) $RMSE_i^p$ $i = 1,2$ denotes the mean of the relative MSE of: (1) the simple average forecast ($RMSE_1^p$), and, (2) the weighted-average forecast ($RMSE_2^p$). In both cases, the MSE of the bias-corrected average forecast is taken as *numeraire*. (B) The superscript p indicates the number of observations in the training sample.

Table 2: Forecast Performance of Philadelphia FED's Survey of Professional Forecasters
Comparing the Simple Average Forecast with the Bias-Corrected Average Forecast

Forecasting Variable	Sample Size	Avg. Bias Estimate $H_0 : B = 0$ (P-Value)	Relative MSE to Feasible Bias-Corr. Avg. Forecast
CPI inflation	$N = 19$ $T = 9$	-0.12 (0.09)	1.09
Unemployment	$N = 22$ $T = 9$	0.032 (0.03)	1.56

Table 3: The Brazilian Central Bank Focus Survey
Computing Average Bias and Testing the No-Bias Hypothesis

Horizon (h)	Avg. Bias $\widehat{B}$	$H_0 : B = 0$ p-value
6	0.075065217	0.09342

Table 4: The Brazilian Central Bank Focus Survey
Comparing the MSE of Simple Average Forecast with that of the Bias-Corrected Average Forecast

Forecast Horizon (h)	(a) MSE Bias-Corr. Avg. Forecast	(b) MSE Avg. Forecast	(b)/(a)
6	0.076	0.085	1.11

Notes: (1) Number of out-of-sample Forecast Periods: 18.